PRM196

VALIDATION OF DISEASE STATES IN SCHIZOPHRENIA: COMPARISON OF CLUSTER ANALYSIS BETWEEN THE UNITED STATES AND EUROPEAN POPULATIONS

Thokagevistk K[1], Millier A[2], Lenert L[3], Sadikhov S[4], Moreno S[4]

[1]*Creativ-Ceutical, Chicago, IL, USA,* [2]*Creativ-Ceutical, Paris, France,* [3]*University of Utah School of Medicine, Salt Lake City, UT, USA,* [4]*F. Hoffmann-La Roche Ltd., Basel, Switzerland*

OBJECTIVES: A set of disease states for patients with schizophrenia was previously published using a statistical clustering method, applied to Positive and Negative Syndrome Scale (PANSS) data from US patients. While factor analyses of the PANSS have shown remarkable stability of the structure across international populations, it is unknown whether similar multidimensional disease states would also be stable. Using data from the European Schizophrenia Cohort (EuroSC), a 2-year observational study in 1,208 schizophrenia patients, we examined the factor structure of the PANSS and identified disease states using the same clustering method as previously. **METHODS:** A principal component analysis (PCA) was conducted using the Kaiser criterion and varimax rotation on PANSS items, followed by a *k*-means cluster analysis on PANSS scores for items most strongly correlated with the PCA domains. For each cluster, a level (low, moderate, high) was assigned to each domain based on the cluster centres values. Kappa statistics were used to measure the agreement in assignment between the published and the derived states sets. **RESULTS:** Five factors accounting for 56% of total variance were obtained from the PCA (positive symptoms, negative symptoms, cognitive impairment, mood disorder, and hostility). As in the analysis of patients in the initial US study, rates of change in root mean squared distance became small after six clusters. When assigning the two sets of states based on levels of positive, negative, and cognitive impairment, the simple, Cicchetti-Allison, and Fleiss-Cohen weighted Kappa statistics (95% CI) were, 0.418 (0.401-0.435), 0.568 (0.553-0.584), and 0.692 (0.676-0.709), respectively. **CONCLUSIONS:** The factor structure, number of discrete states, and combinations of levels of symptoms in states were similar in US and European populations. Resulting moderate-to-substantial agreement in assignment suggests that disease states obtained using *k*-means clustering from the PANSS generalise across international populations.

PRM197

EVALUATING OVERALL SURVIVAL IN ONCOLOGY TRIALS WITH SUBSEQUENT THERAPIES: A METHODOLOGICAL REVIEW AND APPLICATION IN NON-SMALL CELL LUNG CANCER

Jonsson L[1], Fleischer F[2], Bluhmki E[2], Griebsch I[2]

[1]*OptumInsight, Stockholm, Sweden,* [2]*Boehringer Ingelheim Pharma GmbH, Ingelheim am Rhein, Germany*

OBJECTIVES: The use of subsequent therapies has the potential to confound assessment of overall survival (OS) in oncology trials, in particular for trials in early lines of therapy and for malignancies with several registered or investigational treatment options. Standard intent-to-treat analysis is biased, since treatment choices are likely to be influenced by events associated with mortality risk, such as disease progression. We review and compare available statistical methods to obtain unbiased estimates of OS effects in presence of subsequent therapies. **METHODS:** Marginal structural modeling methods include inverse-probability of censoring weighting (IPCW) and inverse-probability of treatment weighting (IPTW). These methods explicitly model both treatment choices and effects of treatments on mortality. Rank-preserving structural failure time models (RPSFT) instead depend on parametric assumptions regarding the effect of investigational and subsequent therapies on survival, and require non-standard estimation methods such as G-estimation or iterative parameter estimation (IPE). We compare the results with the different methods with data from the Lux Lung 1 trial of the tyrosine kinase inhibitor afatinib in non-small cell lung cancer. **RESULTS:** IPCW and IPTW require detailed information on covariates that influence treatment choices and are sensitive to model misspecification. RPSFT may not yield a single estimate of treatment effects due to limitations of the G-estimation procedure. All methods were consistent with a potential OS benefit from afatinib, but the hazard ratio varied from 0.583 (p=0.038) with the pre-specified IPCW method to 0.894 (0.281) with RPSFT/IPE. **CONCLUSIONS:** The proposed methods for obtaining unbiased OS estimates in presence of subsequent therapies rest on assumptions that cannot be tested empirically. There is currently no accepted standard method; pre-specification of model choice is of importance as well as testing alternative methods. Care should be taken to avoid unbalance in subsequent therapy and to record specific information on administered treatments with potential OS effects.

PRM198

QUANTIFYING HEALTH CARE EFFICIENCY: A REVIEW OF PUBLISHED TIME AND MOTION STUDY DESIGN PARAMETERS REFERENCED IN PUBMED BETWEEN 2008-2013

Kritikou P[1], Bassel M[2], De Cock E[3], Payne KA[2]

[1]*Evidera, London, UK,* [2]*Evidera, Dorval, QC, Canada,* [3]*Evidera, Barcelona, Spain*

OBJECTIVES: To review design characteristics of T&M studies applied to health care, with a focus on choice of study design, statistical methodology, and handling of multi-centre data. **METHODS:** A PubMed search was performed using key search terms including "time and motion" (MeSH Term) AND any of the following: cost (analysis), (health) economics, observation(al), and prospective. Articles (English; 2008 or later) were selected based on the following criteria: (1) observational study using T&M methodology; and (2) task-based data collection. Studies that measured broad aggregate health care professional tasks/hospital workflows, in the absence of task- or event-specific timings, were excluded. **RESULTS:** Of 191 identified abstracts, 151 were excluded during screening; upon review, 21 of 40 remaining were retained for detailed assessment. Half (48%) were applicable to Europe, of which 2 were multi-country studies. Medical interventions studied were: drug (48%), diagnostic process (14%), medical procedure (24%), and IT systems to improve clinical management (e.g. EMR) (14%). The majority (86%) of studies were hospital-based, 86% were observational, and 14% employed hybrid methods, including chart review or survey. Only 20% used independent observers. Three quarters (76%) reported descriptive statistics. Of 9 multi-centre studies, one used a random effects regression model to account for "centre clustering", and 8 reported pooled data (3 of which used a "mean of centre averages" approach). Eleven studies (52%) compared two groups, of which 3 applied an analytical design aiming to defect statistical differences, and 2 reported a sample size calculation. **CONCLUSIONS:** This review of T&M studies revealed that descriptive designs are most common (analytical designs using power calculations seem rare). Multi-centre comparator studies rarely use random effects regression models to account for "centre clustering", though considered the method of choice to produce valid confidence intervals around point estimates. In general, statistical methodology is scarcely reported, affecting overall study credibility.

PRM199

THREE TOOLS TO REDUCE THE IMPACT OF COMMON DECISION-MAKING BIASES WHEN CONSIDERING SUBGROUP ANALYSES

Hawkins N[1], Fletcher CA[2]

[1]*Oxford Outcomes Ltd., Oxford, UK,* [2]*Amgen Ltd., Cambridge, UK*

OBJECTIVES: Subgroup analyses of randomized trial data are performed to provide estimates of average treatment effects for patients with specific characteristics. They inform adoption and re-imbursement decisions by identifying groups of patients with favourable risk-benefit or cost-effectiveness ratios. They may also inform decisions regarding the conduct and design of future clinical studies. However, subgroup analyses are essentially observational (patients are not randomized between subgroups) and there is a risk that the differences observed between subgroups may be due to chance rather than reflecting true effects. This risk is exacerbated as typically the same data are used to both select relevant subgroups and to estimate subgroups effects leading to biased estimates and underestimation of uncertainty. This tendency increases as the number of subgroups tested increases. A number of measures are recommended to reduce the risk of bias including: pre-specification, consideration of biological plausibility, and correction of inference for multiple testing. However, the risk of bias is not obviated by pre-specification, correction for multiplicity may lead to discounting of true subgroup effects, and biological plausibility may not be a particularly specific test. In addition, common cognitive and process biases associated with decision-making such as the action imperative, optimism bias, anchoring, and group think may further lead to the inherent uncertainty in subgroup analyses to be effectively underestimated. **METHODS AND RESULTS:** We demonstrate three techniques that may help to counteract these biases: graphical inference methods clearly illustrate the inherent uncertainty in subgroup analysis; Bayesian shrinkage estimation can reduce the effect of anchoring on the observed subgroup effects and encourage consideration of regression to the mean; and reframing exercises (for example, considering the credibility of biological plausibility arguments as if they had been mooted *a priori*) may counter optimism bias. **CONCLUSIONS:** These techniques are illustrated using a published subgroup analysis from the PLATO trial (NCT00391872)

PRM200

PARAMETER IMPORTANCE ASSESSMENT IN A HEALTH ECONOMIC EVALUATION MODEL FOR HEART FAILURE

Cagdas Buyukkaramikli N

Technical University of Eindhoven and Contemporary Analytics, Eindhoven, The Netherlands

OBJECTIVES: Along with uncertainty around the parameters and the initial parameter value assumptions used in health-economic evaluation models, an analysis of the uncertainty around the model inputs/outputs is essential. Parameter importance analysis (PIA) provides an explicit framework to quantitatively identify the contribution of each uncertain input to the output uncertainty. There are several methods available to be used in PIA. The objectives of this research were to investigate different PIA methods with the pros and cons of each method and identify the most robust method with respect to different initial parameter value assumptions. **METHODS:** A health economic model for heart failure is developed to serve as a basis to implement different PIA methods. Six alternative methods are applied: One-way sensitivity analysis, rank correlation analysis, analysis of covariance (ANCOVA), dominance analysis, standardized regression analysis and expected value of perfect parameter information (EVPPI) analysis. Initial parameter assumptions are varied and the robustness of each method is assessed with respect to how close the parameter importance rankings are with different initial parameter assumptions. **RESULTS:** Each technique/initial parameter values' assumption combination generates a different ranking for the importance of the parameters that explain the uncertainty around the expected net monetary benefit with £20,000/QALY. EVPPI is the most robust method with respect to different initial parameter assumptions. However it is the most demanding method in terms of computation time. On the opposite side, one-way sensitivity analysis is the least computation time demanding method; however the importance rankings are very susceptible to change with different initial assumptions. Other Monte-Carlo simulation based methods (e.g. ANCOVA, dominance, standardized regression and rank correlation analysis) are alternative PIA methods, which generate rather robust rankings with different initial parameter assumptions. These alternative methods require substantially less computation times compared to EVPPI with high consistency and robustness to different initial value assumptions.

PRM201

ADJUSTING FOR TREATMENT SWITCHING IN CLINICAL TRIALS WHEN ONLY SUMMARY DATA ARE AVAILABLE – AN EVALUATION OF POTENTIAL METHODS

Boucher R[1], Abrams KR[1], Crowther MJ[1], Lambert PC[1], Wailoo AJ[2], Latimer NR[2]

[1]*University of Leicester, Leicester, UK,* [2]*University of Sheffield, Sheffield, UK*

OBJECTIVES: Treatment switching is an important problem in Health Technology Assessment (HTA), particularly in oncology, which can often bias trial results. Although a variety of statistical approaches have been advocated for adjusting trials subject to treatment switching these all assume that Individual Patient Data

(IPD) is available. In many situations, especially when Indirect Comparison (IC) methods are required to estimate head-to-head effects, it is often the case that IPD is only available for one trial, and summary data for the other. A variety of potential methods are evaluated for the adjustment of such summary data using simulation methodology. **METHODS:** A review of HTA submissions to NICE in which both ICs were used and in which trials were subject to treatment switching was undertaken. A series of simulation studies were undertaken to assess the potential level of bias associated with the methods that are most commonly used for the analysis of such trials. Two broad approaches to adjusting summary data for treatment switching were then evaluated on the simulated data – calculation of Adjustment Factors (AFs), and re-creation and analysis (including bootstrapping) of IPD using scanned survival curves. **RESULTS:** The most commonly reported methods of analysis for studies only presenting summary data were Intention-to-Treat (ITT) and Per Protocol (PP) analyses. Results from the simulation studies indicated that these may be subject to between 0.5% and 140% levels of bias depending on trial characteristics, and that the use of AFs or re-created IPD had potential scope for reducing this. **CONCLUSIONS:** Treatment switching can be associated with considerable levels of bias, and methods for adjusting using summary data, can go some way to compensating for this when IPD is not available as is often the case in Indirect Comparisons (IC). Further extension to a Network Meta-Analysis (NMA) setting is under investigation.

PRM202

MULTIPLE IMPUTATION TECHNIQUES FOR SURVEY DATA WITH MULTIPLE RATING SCALES

Plumpton CO[1], Hughes DA[1], White IR[2]
[1]*Bangor University, Bangor, UK,* [2]*MRC Biostatistics Unit, Cambridge, UK*

OBJECTIVES: Large scale survey data presents a number of challenges to imputation, not least the high number of variables and complexity of the data set. Data may suffer from sparsity in responses, and some questions may be conditional upon previous responses. In addition, survey data commonly contain results from multiple rating scales, which are summed (either directly or weighted) during analysis. We aim to develop a method for the multiple imputation of missing data from complex surveys. **METHODS:** We propose an adaptation of multiple imputation for survey data which contains multiple rating scales, whereby scale summary scores are used within the prediction models. The method is applied to data gathered from a large multinational survey, with data sets from 9 countries. Analysis uses a logistic regression model on each of the 9 data sets, and results are compared from a complete case analysis approach with those from multiple imputation. **RESULTS:** The proposed approach reduces the size of the prediction models from 135 predictors to a maximum of 72. Distributions of imputed data are seen to be consistent with observed data. Results from the regression analysis with multiple imputation are similar to, but show lower standard errors than, results for complete case analysis; for the same regression models a 39% reduction in the standard error is observed. **CONCLUSIONS:** Our adaptation makes multiple imputation practical for large scale survey data with multiple rating scales. For the data considered, analysis of the multiply imputed data shows greater power and efficiency than complete case analysis. The adaptation of multiple imputation makes better use of available data and can yield substantively different results from simpler, less valid techniques.

PRM203

STRUCTURAL FAILURE TIME MODELING OF OVERALL SURVIVAL EFFECTS IN ONCOLOGY TRIALS WITH SUBSEQUENT THERAPIES

Jonsson L[1], Fleischer F[2], Bluhmki E[2], Griebsch I[2]
[1]*OptumInsight, Stockholm, Sweden,* [2]*Boehringer Ingelheim Pharma GmbH, Ingelheim am Rhein, Germany*

OBJECTIVES: Subsequent therapies can confound the evaluation of overall survival (OS) in oncology trials. We evaluated the application of rank-preserving structural failure time modeling for the estimation of OS effects in presence of subsequent therapies through Monte Carlo simulations. Results were demonstrated for a clinical trial: the Lux Lung 1 study of afatinib vs. placebo in non-small cell lung cancer. **METHODS:** In accelerated failure time models, covariates are assumed to affect survival times rather than hazard rates. Counterfactual survival times can therefore be computed, i.e. how long patients would have survived without the investigational or subsequent therapies. The parameters of structural failure time models can be obtained by G-estimation, whereby counterfactual survival times are calculated with hypothetical treatment effects and OS is compared between treatment arms. The G-estimate is the set of hypothetical effects that generate the most similar survival in both study arms. Branson & Whitehead (2002) developed an alternative estimation method for trials with cross-in from placebo to active treatment based on iterative parametric regressions; we extend this framework to the application with subsequent therapies. **RESULTS:** Simulation showed that standard methods are biased in the presence of subsequent therapies affecting overall survival. This includes intent-to-treat analysis, censoring at start of subsequent therapies and subgroup analysis in patients never receiving subsequent therapy. G-estimation often failed to identify parameter values when more than one treatment effect was included in the model. Iterative parameter estimation produced unbiased estimates in simulation studies and predicted a small numeric but non-significant survival benefit of afatinib. **CONCLUSIONS:** Structural failure time models can be useful to obtain unbiased estimates of OS in presence of subsequent therapies. However the assumption of proportionality in survival times cannot be tested empirically and non-standard estimation procedures are required.

PRM204

THERE IS MORE TO DECISION MAKING THAN COSTS AND EFFECTS: HANDLING PRACTICAL CONSTRAINTS IN THE VALUE OF INFORMATION FRAMEWORK

Koffijberg H, Janssen M
University Medical Center Utrecht, Utrecht, The Netherlands

OBJECTIVES: Whether new medical technology is implemented may depend on the balance between costs and effects, but also on practical constraints. Examples are a fixed health care budget and a maximum clinically acceptable risk of adverse events. However, the impact of compliance with such constraints cannot be handled explicitly in the current value of information (VOI) framework. Our objective was to demonstrate proper handling of constraints by extending the VOI framework through separation of cost, effect, and constraint components. **METHODS:** The proposed VOI extension was investigated in a simulation study comparing two hypothetical drugs and their side effects. The VOI extension was also applied to a clinical study concerning the cost-effectiveness of carotid intima-media thickness measurements to improve treatment guidance of patients at high risk of cardiovascular disease. Results of the standard VOI analysis, considering only costs and effects, were compared with results from the extended VOI analysis explicitly considering constraints. **RESULTS:** Standard VOI results may under- or overestimate the value of additional research compared to extended VOI results. In our clinical example, with penalties of $2 and $5 per dollar budget exceedance, standard values for the Expected Value of Perfect Information (EVPI) of $24, and $1,490 were found, with corresponding values of $239, and $565 for the extended EVPI. Ignoring the budget constraint in the standard EVPI analysis therefore resulted in a underestimation of $214 ($2 penalty) and an overestimation of $925 ($5 penalty) of the EVPI per patient. **CONCLUSIONS:** When decision-maker's criteria go beyond costs and effects, standard VOI results may not reflect the actual value of additional research accurately and may therefore jeopardize optimal research prioritization. Determination of the extended VOI, through separation of cost, effect, and constraint components, is straightforward and can support optimal research prioritization regardless of the complexity of the decision criteria considered.

PRM205

MEASURING TREATMENT EFFECTS ON RARE EVENTS USING META-ANALYSIS: AN ASSESSMENT OF EXISTING METHODS

Jamotte A, Roïz J
Creativ-Ceutical, London, UK

OBJECTIVES: Meta-analysis combines results from independent studies to produce robust statistical estimates. This technique is widely used in health care to synthesise treatment effects from clinical studies. However, when dealing with rare events such as rare adverse events, existing meta-analysis methods might not produce good treatment effect estimates, especially when there is no event occurrence in one or both arms of a study. The objective of this study is to compare the performance of various methods in estimating effect size for rare events. **METHODS:** An assessment of meta-analysis methods providing pooled odds-ratios as effect size estimates was conducted for different scenarios. The Inverse Variance Weighted, Peto, Mantel-Haenszel and logistic methods were assessed, with constant, "treatment arm" or empirical continuity corrections added when needed. The scenarios were created using different values of odds-ratio, baseline risk, and group imbalance. For each scenario 5,000 simulations of 10 studies were generated using R software. Coverage, bias and statistical power were used to compare the methods. **RESULTS:** The most commonly used continuity correction is outperformed in every scenario by the two other corrections. The inverse variance method, most commonly used in meta-analysis, performs poorly when the event probability is smaller than 0.10: it is not recommended for sparse data. Peto's method performs well in some scenarios but leads to biased results with high odds ratios and high imbalance. The logistic method is highly biased when baseline risk is low and true odds ratio is high. Under other scenarios it performs well but is most often outperformed by other methods. The Mantel-Haenszel method with empirical correction performs constantly well over the scenarios. **CONCLUSIONS:** These findings may be used to develop guidelines on when to use which method for conducting meta-analysis with rare events. Next steps will be to assess the use of mixed models and Bayesian techniques.

PRM206

METHODOLOGY FOR ESTABLISHING INTERNAL AND EXTERNAL VALIDITY WHEN PROPENSITY SCORE MATCHING IS USED

Eisenberg D[1], Wasser T[1], Placzek H[2], Luthra R[1]
[1]*HealthCore, Inc., Wilmington, DE, USA,* [2]*HealthCore, Inc., Andover, MA, USA*

OBJECTIVES: Propensity score matching (PSM) is an approach commonly used when treatment and control groups are thought to be different on key study variables. When the control group is larger than the treatment group, (as large as 20:1) a good match might be easy to obtain. However, differences may exist between the matched controls and the unmatched controls, indicating poor generalizability of study results. **METHODS:** Groups for the analysis are the unmatched controls (UM), the matched controls (MC) and the treatment cohort (TRT). Analysis methods for these groups in a fully crossed method and interpretation of the results will determine internal (IV) and external validity (EV). Analysis comparing the groups against the outcomes variable will determine if variables need to be controlled for in models that may be developed. **RESULTS:** After the PSM is conducted MC and TRT groups should be compared on the matched variables. Differences at this stage would indicate a poor match and a low level of IV. MC and UM should also be compared on the variables used for matching, as well as the outcome variables of interest. Significant differences on the matched variables would indicate low EV and poor generalizability of results, while differences of MC and UM groups and UM and TRT groups on the outcome variables would indicate that statistical models would need to address covariates as potential confounding effects would be present. Analysis methods can be fit statistics (chi-square or equivalence tests) or typical inferential methods with adjusted p-values greater than 0.05. **CONCLUSIONS:** It is important that research studies maintain good IV and EV. This is often complicated in research where the controls vastly outnumber the treatment group. Proper statistical analysis can go a long way to test and clarify data to make the results as meaningful as possible.